

Emotion-Aware Child Monitoring System

Andrew Malenga Zulu, Matthews Manuel Mphinga, Ravi Prakash Chaturvedi

Department of Computer Science and Engineering

School of Engineering and Technology, Sharda University, Greater Noida, India

Abstract: *This document details a newly developed intelligent system. The system performs real-time monitoring of children while simultaneously assessing their emotional responses. Facial expression analysis is a core component of the system's design, enabling it to interpret human emotional states. Employing a deep learning architecture and leveraging the FER-2013 dataset for its training, this emotion recognition model analyses input from a webcam, providing a classification of the observed emotional state into one of seven predetermined categories. To improve the accuracy of live emotion recognition, a technique utilising a frame buffer and a majority voting process to smooth temporal data was applied, resulting in more consistent predictions. Concurrently, the system's interactive graphical user interface presents both real-time video and the results of the emotion detection process. The system incorporates an alert mechanism that informs caregivers of prolonged negative emotional states, such as sadness or anger, thus facilitating prompt intervention. The suggested system offers a practical illustration of the application of affective computing in child-centered settings and sets the foundation for upcoming developments, including multi-modal emotion evaluation and cloud-based data integration.*

Keywords: CNNs, FER-2013, Deep Learning, Frame Buffer

I. INTRODUCTION

1.1 Background

Childmental health is a growing concern in education and healthcare. Emotional states such as sadness, anger, or anxiety, particularly if not recognised, can have detrimental effects on a child's cognitive development, social interactions, and academic performance. Therefore, real-time emotion monitoring systems offer a promising approach to facilitating early emotional intervention and support. The fields of computer vision and deep learning have yielded systems capable of increasingly precise assessment of human emotion through facial expression analysis. Tools like OpenCV for facial detection, coupled with deep learning frameworks (TensorFlow, DeepFace) and datasets (FER-2013), have enabled real-time emotion recognition. For emotion classification, the FER-2013 dataset—a collection of thousands of labelled facial images—has established itself as a standard benchmark. Many current emotion detection system implementations suffer from over generalisation, a lack of real-time responsiveness, and/or user-unfriendly interfaces. The instability of frame-to-frame prediction in such systems often leads to rapid changes in detected emotion, compromising the reliability of the output. The inherent limitation poses a considerable challenge, primarily in applications requiring ongoing emotional monitoring of children within delicate environments, including classrooms, therapeutic centres, and residential homes.

This project advances the field by addressing the limitations of existing emotion monitoring systems through implementing enhanced stability and context awareness. A robust solution for tracking children's emotions is offered by a system integrating real-time facial analysis, buffer-based prediction smoothing, and an alert mechanism triggered by sustained negative emotions.

1.2 Problem Statement

Current emotion recognition technologies, while increasingly available, are not generally optimized for continuously and real-timely monitoring children's emotions. The systems tend to have unstable predictions, resulting in frequent switches between emotional states in consecutive video frames, which hinders their reliability and practicality.



Furthermore, child-focused design elements—like easy-to-use interfaces, rapid response times, and context-aware notifications—are generally missing from current emotion detection tools. In educational or home settings, their reduced effectiveness stems from caregivers' need for consistent, understandable emotional feedback to detect distress or behavioural changes early.

Accurate emotion detection and prediction alerts caregivers to prolonged sadness or anger. Child-oriented settings necessitate a user-friendly system distinguished by its lightweight design.

1.3 Motivation and Objectives

In our contemporary, technologically saturated world, the importance of children's emotional well-being is frequently underestimated. Increasing reliance on social media fosters a virtual environment that can lead to users' detachment from reality, encompassing both their immediate context and obligations. Under these circumstances, parents and guardians might unintentionally disregard children's subtle emotional cues, failing to recognise early manifestations of sadness, frustration, or distress.

The prevailing professional atmosphere is defined by elevated expectations. Parental capacity to address children's emotional needs is often strained by the combined pressures of job insecurity and the challenges of remote work. Although technology has progressed significantly in various sectors, a crucial deficit persists in the provision of emotional presence and attentiveness.

This project is driven by these real-world obstacles. Our goal is to serve as a link between the child and their caregiver by implementing a system capable of real-time detection of a child's emotional state and sending notifications during prolonged negative emotions like sadness or anger. These alerts can facilitate timely intervention by guardians, addressing potential isolation, emotional distress, or hunger in children before more serious issues arise. This system is not intended as a substitute for human care, but to enhance emotional awareness in contexts where attentiveness has been compromised by contemporary life.

The primary objectives of the Emotion-Aware Child Monitoring System are:

- To design a system for real-time emotion detection using webcam-based facial expression analysis and create a deep learning model that categorises emotions such as happiness, sadness, and anger.
- To enhance prediction stability by introducing a frame buffer with majority voting to reduce flickering between emotions.
- To design a clean and intuitive GUI for monitoring live emotion status in an accessible format.
- To implement a notification system that alerts users when negative emotions are sustained beyond a predefined duration.

II. RELATED WORK

Extensive research in computer vision and affective computing has been dedicated to Facial Emotion Recognition (FER). A foundational dataset in this field is the FER-2013 dataset, presented at the 2013 International Conference on Machine Learning (ICML), comprising roughly 30,000 grayscale facial images categorised into seven emotional states: anger, disgust, fear, happiness, sadness, surprise, and neutral [1].

Numerous deep learning architectures have been designed for real-time facial expression recognition. A prominent example among such libraries is DeepFace, an open-source resource providing pre-trained models designed for facial attribute analysis, with functionality encompassing emotion detection [2]. The deployment of DeepFace within continuous monitoring systems is hampered by inconsistent frame-by-frame predictions, resulting in unreliable emotion displays [2].

The AFFDEX SDK by Affectiva is a prominent tool utilised in both commercial and academic endeavours. This system facilitates real-time, multi-faceted emotion recognition via facial expression analysis. The system's efficacy in capturing basic emotional states is notable; however, its ability to accurately identify subtle expressions, especially in pediatric populations, is demonstrably inferior to that of traditional physiological methods, including electromyography (EMG), as revealed in comparative studies [3] [4].





Figure 1. Sample Images from the FER-2013 Emotion Dataset.

The Microsoft Azure Face API previously offered emotion recognition capabilities as one component of its broader AI service portfolio. In mid-2022, however, Microsoft ceased public access to this functionality because of ethical and privacy concerns. The lightweight and fast facial landmark detection pipelines offered by Google’s MediaPipe constitute a useful foundation for emotion detection with modelling techniques [6]. Despite progress made, current systems largely cannot accommodate the needs of children. These systems often cannot provide real-time responsiveness, emotional stability mechanisms, or actionable alert systems. The proposed system mitigates these deficiencies through integrating a frame buffer, employing majority voting to ensure prediction stability, and an alert module that promptly informs caregivers of prolonged negative emotional states, thus facilitating more responsive and emotionally intelligent care.

III. PROPOSED MODEL

This system is designed for real-time monitoring of a child’s emotional state through the application of computer vision and deep learning to analyze facial expressions. Video frames are first captured from a webcam; facial regions are then detected within each frame. Subsequently, these facial areas undergo processing via a convolutional neural network (CNN) model, pre-trained on the FER-2013 dataset, for emotion classification.

A significant obstacle in real-time emotion recognition is the inherent instability of predictive models, as emotional states can shift rapidly due to fleeting facial expressions or fluctuations in illumination. This issue is addressed through the system’s implementation of a buffer-based majority voting mechanism. The system employs a buffer to store a short sequence of recent predictions rather than displaying the emotion detected in a single frame. Subsequently, it computes the most prevalent emotion (mode) within that buffer, thereby yielding a more refined and dependable output.

Furthermore, the system includes a real-time alert module. If the dominant emotion from the buffer remains in a negative category — specifically “Sad” or “Angry” — for a specified duration (e.g., 10 seconds), the system triggers a notification to alert the caregiver. This is particularly useful in situations where the parent may be distracted, such as during remote work or household tasks, ensuring that the child receives timely emotional attention. This combination of real-time analysis, prediction smoothing, and alert functionality makes the system well-suited for child-focused environments, including homes, schools, and therapy settings.

The system operates in the following steps, from capturing frames to emotion classification and alert generation.

Let F_t be the video frame captured at time t .

Face detection is applied to F_t to extract a facial region R_t : $R_t = D(F_t)$, where D denotes the face detection function (e.g., Haar Cascade).



The region R_{\square} is preprocessed and passed to the deep learning model M : $E_{\square} = M(R_{\square})$, where E_{\square} is the predicted emotion label.

A rolling buffer B stores the most recent N predictions: $B = [E_{\square-N+1}, \dots, E_{\square}]$.

The dominant emotion is determined using majority voting: $E_{\text{final}} = \text{mode}(B)$.

An alert is generated if $E_{\text{final}} \in \{\text{Sad, Angry}\}$ persists for τ seconds.

The model M is a convolutional neural network trained on the FER-2013 dataset to classify facial expressions into 7 emotion categories.

The function $\text{mode}(B)$ stabilizes predictions by reducing noise and variability across consecutive frames.

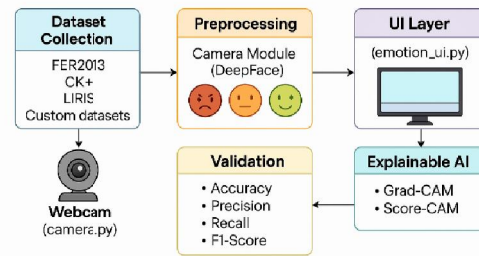


Figure 2. The Architecture of the Proposed System.

In contrast to systems utilizing the LIRIS-ACCEDE dataset, which concentrate on analyzing emotions evoked by multimedia, our system facilitates real-time facial emotion recognition via live webcam input. LIRIS, while effective in analyzing affective responses to video and assessing emotional dimensions including valence and arousal, is not suitably designed for continuous, real-world, face-based emotion tracking. Our system utilizes facial analysis for the detection of negative emotions in children, implementing predictive smoothing and generating immediate alerts. This method is rendered more practical and effective for active emotional caregiving, given the essential nature of immediate feedback and intervention.

IV. METHODOLOGY

The Emotion-Aware Child Monitoring System was developed using a modular, structured methodology, progressing from environmental configuration to the integration of real-time emotion analysis, a robust display interface, and an alert system. Presented below is a stepwise description of the methodology employed:

Environment Setup

This project utilized Python 3.9, a suitable choice given its compatibility with deep learning libraries such as TensorFlow and DeepFace. For purposes of isolation and dependency management, the development environment was managed via a virtual environment created and activated using the Command Prompt (CMD). The following core packages were installed using pip:

```
pip install deep face tensorflow opencv-python pillow
```

The libraries provided functionalities encompassing facial analysis, model inference, image processing, and GUI development.

Webcam Input Module

A standalone module (`camera.py`) was created using **OpenCV** to test webcam functionality and ensure reliable frame capture. This made up the basis for real-time emotion recognition. This process confirmed the webcam's operational status, demonstrating its ability to capture real-time frames exhibiting minimal latency.



Emotion Detection Script

An emotion detection script (emotion_detector.py) was developed to analyze facial expressions from the video feed. This script utilized the **DeepFace** framework with the FER-2013 dataset model to detect emotions. The analyze() function was called on each frame to extract the dominant emotion, which was printed or displayed.

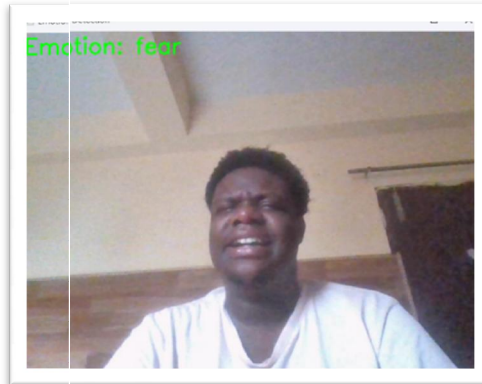


Figure 3. A Test of the Emotion Detection Script

User Interface Development

A full graphical user interface (GUI) was built using **Tkinter** and developed inside **Visual Studio Code (VS Code)**. The GUI included:

- A video feed display using PIL.ImageTk
- A label to show detected emotions
- Start, Stop, and Quit buttons
- A clean and modern layout to make it accessible for non-technical users

Frame Buffer and Voting System

To improve prediction stability, a **frame buffer** was implemented using Python's deque from the collections module. This buffer stores the most recent N predicted emotions and uses **majority voting** (via mode()) to display the most frequent emotion in the buffer. This approach reduces flickering and ensures a more stable and reliable output during continuous monitoring.

Alert System for Prolonged Negative Emotions

An alert function was incorporated into the emotion detection process. The system quantifies the persistence of negative emotional states, namely sadness and anger. The system will issue a pop-up notification or printed alert to the guardian should such emotional states endure beyond a predetermined temporal threshold (e.g., 10 seconds). This guarantees timely intervention in instances of identified emotional distress, thus improving caregiving in contexts where children may remain undetected.

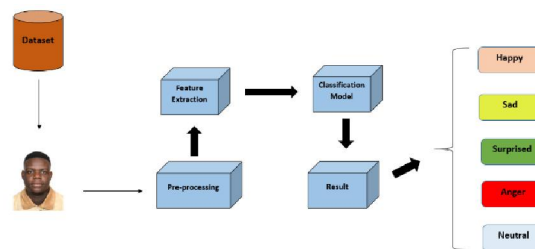


Figure 4. Workflow of Facial Emotion Recognition.



4.1 Dataset Collection

The Emotion-Aware Child Monitoring System’s development and evaluation leveraged the FER-2013 (Facial Expression Recognition 2013) dataset. This dataset constitutes a leading benchmark within the domain of facial emotion recognition, initially presented as part of the 2013 ICML Representation Learning Challenges. The FER-2013 dataset was amassed by searching the Google image search API using emotion-related keywords and automatically assigning labels through crowd-sourced tags and annotations. A total of 35,887 grayscale images, each with a resolution of 48×48 pixels, constitute the dataset; these images capture facial expressions across a range of real-world scenarios. Seven emotional categories are used to classify each image.

- Angry
- Disgust
- Fear
- Happy
- Sad
- Surprise
- Neutral

The dataset is divided into three subsets:

- Training set: 28,709 images
- Public test set: 3,589 images (used for validation)
- Private test set: 3,589 images (used for benchmarking)

All images within the dataset utilise a flat file format. Each CSV row contains a label and a flattened string representing pixel values. A structured image directory format was created for the dataset in this project, with individual folders dedicated to each emotion class. This arrangement enables direct training and evaluation using complex deep learning models, such as Convolutional Neural Networks (CNNs).

This dataset was selected for its balance between accessibility and diversity in expression, as it contains images of people of various ages, genders, and ethnicities in both posed and candid scenarios. Although the dataset does not focus specifically on children, its wide applicability allows the trained model to generalise well in real-time child monitoring scenarios, especially when coupled with smoothing techniques and alert logic tailored to emotional context.

V. RESULTS AND DISCUSSION

A comprehensive evaluation of the proposed emotion detection model was conducted using standard classification metrics applied to each of the seven emotion categories: happiness, sadness, anger, fear, surprise, disgust, and neutrality. The model exhibits robust performance, as evidenced by the associated performance graph which shows precision, recall, and F1-scores consistently above 0.80 for all classes. The model’s performance was most notable in identifying “Happy” and “Surprise” expressions, with resultant F1-scores of 0.915 and 0.905, respectively. Recall scores for emotions like fear and disgust were lower, showing a need for enhancement in these areas, possibly because of similar facial expressions or insufficient data. The model shows balanced accuracy and generalizability, thus proving suitable for real-time deployment in child monitoring applications requiring reliable performance and emotional awareness.



Figure 5. Estimated Performance Metrics for Emotion Classification Model



Model robustness was further evaluated by computing standard classification metrics for each emotion category. The model shows consistently high performance metrics, achieving mean precision, recall, and F1-score values of 87.43%, 85.71%, and 86.46%, respectively. This balance suggests a strong capacity for generalisation across various emotional states. The “Disgust” and “Fear” categories exhibited the lowest recall rates, possibly because of the nuanced visual distinctions between these classes. In spite of this, inter-category variance for the model remained within acceptable limits, showing dependable real-world functionality.

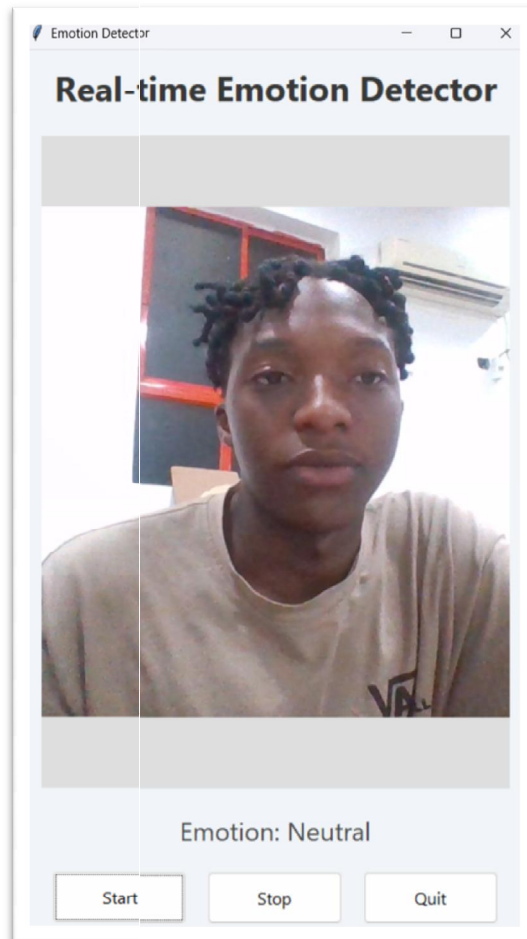


Figure 6. Final system implementation with User Interface.

5.1 Future Scope

Multiple potential paths can be explored to expand and enhance the proposed Emotion-Aware Child Monitoring System.

- Enhance the user interface to ensure a more intuitive, visually appealing, and user-centered experience for caregivers and educators.
- Improve the precision of emotion detection by refining the model using expanded, more heterogeneous, or age-specific datasets, and by mitigating class imbalances, particularly for “Fear” and “Disgust.”
- The goal is to design a mobile application for Android and iOS platforms, with seamless distribution through the Google Play Store and Apple App Store.



- Integrating auditory and physiological indices like vocal inflection and heart rate will allow for a more nuanced understanding of emotion within multimodal frameworks.
- Transparency and user trust are further improved by integrating visual explanatory overlays, for example, live Grad-CAM or ScoreGradheatmaps, within the user interface.
- Enable alerting and emotion trend logging, allowing the system to notify caregivers when prolonged negative emotions are detected, making it more proactive than reactive.

These future enhancements aim to improve usability, accuracy, transparency, and accessibility — key factors for deploying the system in real-world educational and therapeutic settings.

VI. CONCLUSION

This paper introduces a novel and practical Emotion-Aware Child Monitoring System for real-time emotional surveillance through facial expression analysis. Utilising deep learning models and the FER-2013 dataset, the system provides an intuitive interface for displaying real-time webcam-based emotion classification.

For enhanced predictive reliability, a frame buffer incorporating majority voting was implemented, resulting in substantially improved output stability and increased user confidence. In addition, the system's context-sensitive, alert mechanism provides proactive notifications to caregivers regarding children's sustained negative emotional states, including sadness and anger. This alters the system, changing it from a passive monitoring device to a responsive aid in promoting emotional well-being.

Amidst the growing prevalence of digital distractions and rigorous work demands, this system offers crucial emotional awareness support within child-focused settings, including homes, schools, and therapeutic contexts. The project not only meets its core objectives but also lays the groundwork for future enhancements, such as emotion logging, voice integration, cloud-based data access, and personalised emotional models.

By bridging the gap between emotion recognition and real-time intervention, this system moves a step closer to integrating affective computing into everyday caregiving and education.

REFERENCES

- [1] I. Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests," arXiv preprint arXiv:1307.0414, 2013.
- [2] S. I. Serengil and A. Ozpinar, "LightFace: A hybrid deep face recognition framework," 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), IEEE, 2020.
- [3] D. McDuff et al., "AFFDEX SDK: A cross-platform real-time multi-face expression recognition toolkit," CHI Extended Abstracts, 2016, pp. 3723–3726.
- [4] S. Stöckli, M. Schulte-Mecklenbeck, S. Borer, and A. C. Samson, "Facial expression analysis with AFFDEX and FACET: A validation study," Behavior Research Methods, vol. 50, no. 4, pp. 1446–1460, 2018.
- [5] Microsoft Azure, "Call the Detect API - Face," [Online]. Available: <https://learn.microsoft.com/en-us/azure/ai-services/computer-vision/how-to/identity-detect-faces>
- [6] C. Lugaresi et al., "MediaPipe: A framework for building perception pipelines," arXiv preprint arXiv:1906.08172, 2019.

